

DOI: 10.19666/j.rlfed.202305073

# 基于机器学习的飞灰含碳量预测模型 比较研究

陈植元<sup>1</sup>, 谭厚章<sup>2</sup>, 成思扬<sup>3</sup>, 张诗雪<sup>1</sup>, 熊小鹤<sup>2</sup>, 阮仁晖<sup>2</sup>

(1. 武汉大学经济与管理学院, 湖北 武汉 430072;

2. 西安交通大学热流科学与工程教育部重点实验室, 陕西 西安 710049;

3. 佐治亚理工学院, 佐治亚州 亚特兰大 30332)

**[摘要]** 锅炉飞灰含碳量是衡量锅炉燃烧效率的重要指标之一, 基于机器学习构建了一套能够准确预测飞灰含碳的模型。首先, 借助随机森林算法解决了飞灰含碳实测值与其他输入特征频率不一致的问题; 其次, 采用基于随机森林的递归特征消除方法, 从30个原始输入特征中提取出9个输入特征, 在降低模型计算量的同时提高了预测准确度; 最后, 以某电厂330 MW 机组锅炉实际运行数据, 建立了线性回归、决策树、KNN、随机森林、Catboost、XGBoost 6个机器学习模型对飞灰含碳量进行预测。预测结果发现: 决策树、KNN、随机森林和XGBoost模型预测效果较好, 均方误差分别为0.010、0.009、0.006和0.006, 线性回归模型表现最差; 构建的预测模型在锅炉低、中、高负荷下均保持稳定。

**[关键词]** 飞灰含碳量; 随机森林; XGBoost; 递归特征消除

**[引用本文格式]** 陈植元, 谭厚章, 成思扬, 等. 基于机器学习的飞灰含碳量预测模型比较研究[J]. 热力发电, 2023, 52(7): 64-73. CHEN Zhiyuan, TAN Houzhang, CHENG Siyang, et al. Comparison of prediction models of carbon content of fly ash based on machine learning[J]. Thermal Power Generation, 2023, 52(7): 64-73.

## Comparison of prediction models of carbon content of fly ash based on machine learning

CHEN Zhiyuan<sup>1</sup>, TAN Houzhang<sup>2</sup>, CHENG Siyang<sup>3</sup>, ZHANG Shixue<sup>1</sup>,  
XIONG Xiaohe<sup>2</sup>, RUAN Renhui<sup>2</sup>

(1. School of Economics and Management, Wuhan University, Wuhan 430072, China;

2. MOE Key Laboratory of Thermo-Fluid Science and Engineering, Xi'an Jiaotong University, Xi'an 710049, China;

3. Georgia Institute of Technology, Atlanta, Georgia 30332, America)

**Abstract:** The carbon content of fly ash in boilers is one of the important indicators of combustion efficiency. This study employs machine learning models to accurately predict the carbon content of fly ash. Firstly, random forest is employed to adjust the frequency of fly ash carbon content data to once per minute, aligning it with the input features to address the issue of imbalanced data collection frequency. Then, a recursive feature elimination method based on random forest is used to extract nine important features out of the original 30 features, reducing feature correlation and improving model accuracy. Subsequently, six machine learning models (linear regression, decision tree, K-nearest neighbors (KNN), random forest, Catboost and XGBoost) are compared for prediction. The results indicate that decision tree, KNN, random forest and XGBoost models perform well, MSE of which on the test are 0.010, 0.009, 0.006 and 0.006, respectively, while linear regression exhibits the poorest performance. The prediction models remain robust under low, medium, and high boiler loads.

**Key words:** carbon content of fly ash; random forest; XGBoost; recursive feature elimination

收稿日期: 2023-05-24

基金项目: 国家自然科学基金项目(71871166)

Supported by: National Natural Science Foundation of China (71871166)

第一作者简介: 陈植元(1983), 男, 博士, 副教授, 博士生导师, 主要研究方向为运营管理, 机器学习, zhiyuanchen@whu.edu.cn.

锅炉的飞灰含碳量是煤粉在锅炉燃烧后剩余飞灰中碳的质量分数。当煤粉在锅炉内燃烧不充分时,飞灰含碳量会增大,未燃尽损失增大,锅炉效率降低,发电煤耗升高,其是反映火力发电厂锅炉燃烧效率的重要指标之一。准确预测飞灰含碳量有利于发电厂及时调整设备从而提高锅炉燃烧水平,降低发电成本。

目前,对于飞灰含碳量的测量方法多采用灼烧失重法。灼烧失重法虽然测量精度高,但操作十分繁琐,需要经过采样、制样、灼烧和称重等步骤。该方法存在较长的滞后时间,无法及时指导机组运行人员对锅炉的各种参数进行调整。因此部分人员开始摒弃该方法,而采用机器学习方法进行预测。机器学习方法的基本思想是通过混合方法确定影响飞灰含碳量的重要输入特征,然后运用这些特征建立机器学习模型预测飞灰含碳量。与其他方法相比,该方法测量成本低,并且预测相对准确。

关于预测飞灰含碳量的机器学习模型,王春林等<sup>[1]</sup>提出了基于支持向量机的预测模型,并通过热态实验对该模型的有效性进行了验证。陈敏生等<sup>[2]</sup>基于核主元分析和支持向量机建立了飞灰含碳量预测模型。周国雄等<sup>[3]</sup>利用变学习率 BP 神经网络、主成分分析、在线支持向量机、改进灰色预测等方法建立了基于信息熵的智能集成预测模型。卞和营等<sup>[4]</sup>建立了基于支持向量回归预测模型,并通过交叉验证法优化了惩罚参数和核函数参数。周昊<sup>[5]</sup>、方湘涛<sup>[6]</sup>、赵新木<sup>[7]</sup>、温文杰<sup>[8]</sup>、叶兆青<sup>[9]</sup>、冯旭刚<sup>[10]</sup>、乔源<sup>[11]</sup>等利用 BP 神经网络构建了飞灰含碳量预测模型。王芳等<sup>[12]</sup>根据变量重要性评分选择变量,建立了基于随机森林的预测模型。朱璘琦等<sup>[13]</sup>将主成分分析法或 Garson 算法与普通 LM-BP 神经网络相结合。张先练等<sup>[14]</sup>提出分层多项式回归预测模型。但是,目前预测飞灰含碳量的机器学习方法仍存在以下问题:1)可采取的锅炉输入特征众多,但部分输入特征可能对飞灰含碳量影响不大。同时许多特征间的相关性较强易导致多重共线性,增加了模型计算复杂度;2)目前研究主要集中于单一的机器学习模型,如随机森林或神经网络等,缺乏较为全面的多模型预测对比;3)飞灰含碳真实数据与输入特征的数据采样频率不匹配。电厂每 8 h 取一次飞灰进行含碳量化验,只能获得 8 h 内的平均值,但锅炉的输入特征是每分钟采样一次,解决飞灰含碳量与输入特征数据采样频率不一致是亟需解决的问题。

基于以上问题,本文选取某电厂 330 MW 机组锅炉 9 天的实际运行数据,基于随机森林模型的递归特征消除方法得到影响飞灰含碳量的主要输入特征,并建立 6 种机器学习模型,通过多种评价指标对模型预测效果进行评估分析,获得表现较好的预测模型。

## 1 建模理论

主要机器学习模型的优缺点见表 1。采用递归特征消除(recursive feature elimination, RFE)方法提取特征,RFE 通过反复迭代,每次迭代中训练模型并剔除最不重要的特征,最终得到一个有限数量的重要特征子集。该方法可以自动选择最优特征子集,减少特征数量和模型复杂度,提高模型的泛化能力和计算效率。基于随机森林模型的 RFE 是将需要的特征集初始化为整个数据集,每次剔除一个排序准则分数最小的特征,直到获得最后的特征集,该算法是一个基于随机森林模型的最大间隔原理的序列后向选择算法<sup>[15]</sup>。

表 1 机器学习模型  
Tab.1 Machine learning models

模型名称	模型优点	模型缺点
线性回归	简单、易于理解和实现,计算效率高。对于线性关系较强的数据集表现良好	对非线性数据拟合较差
决策树	易实现;易解释;无需对数据有任何的先验假设;稳健性强	容易过拟合,对于数据的小变化敏感,不稳定性较高
KNN	简单,对异常值不敏感	对稀有类别样本的预测准确率低;解释性弱
随机森林	能处理高维数据,具有较好的准确性和鲁棒性	模型可解释性较差,对于噪声较多的数据容易过拟合,需要较多的计算资源
Catboost	对分类变量的处理能力强,能够处理大量特征和样本,具有较好的准确性和泛化能力	需要大量内存和时间处理类别型变量;随机数的选择影响模型预测结果
XGBoost	处理稀疏数据和高维特征表现良好,具有较高的准确性和鲁棒性	需要调整的参数过多

## 2 对象描述

研究锅炉为亚临界、中间一次再热、控制循环汽包炉,燃料为晋中贫煤,采用钢球磨,中间仓储制粉,热风送粉,四角切圆燃烧。主燃烧器结构包括 4 层一次风喷口、7 层二次风喷口(其中 1 层二次风喷口常开)、2 层三次风喷口、1 层 OFA 喷口及 4 层 SOFA 喷口,锅炉与燃烧器组结构如图 1 所示。

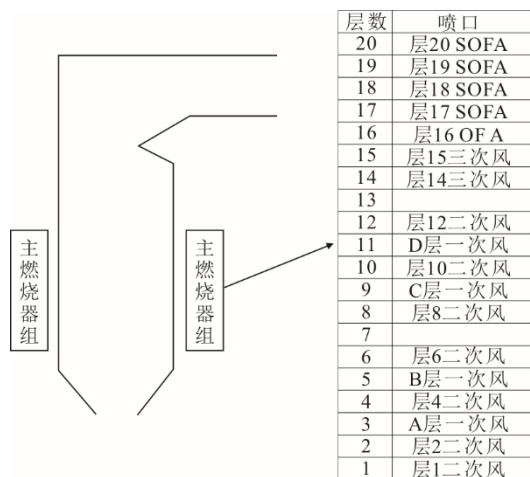


图1 锅炉与燃烧器组结构示意图  
Fig.1 Boiler and burner suit structure

本文数据采样区间为2022年7月22日至2021年7月30日,共含12 000条观测数据,横跨早中晚、低中高负荷等不同工况,具有一定代表性。其中飞灰含碳量为每8 h取1次,并有A、B两侧数据,锅炉运行参数为每1 min取1次。通过对锅炉燃烧机理的分析,36个运行参数被选取作为模型的输入特征,具体见表2。在建模中将数据分为A侧、B侧和AB侧平均3种情况分别进行,A侧和B侧的初始特征各有31个,下文数据处理与建模均以A侧数据示例。

表2 模型输入特征  
Tab.2 Input features of the model

特征名称	特征名称
机组负荷	层17 SOFA 挡板
主蒸汽流量	层18 SOFA 挡板
总风量	层19 SOFA 挡板
一次风管与炉膛差压	层20 SOFA 挡板
燃烧器摆角指令信号 A	层5 周界风挡板
炉膛烟气出口温度1	层9 周界风挡板
炉膛烟气出口温度2	层11 周界风挡板
层3 给粉机转速之和	层16 OFA 挡板开度
层5 给粉机转速之和	A侧二次风箱与炉膛差压
层9 给粉机转速之和	A侧二次风箱入口压力
层11 给粉机转速之和	A侧炉膛出口温度
层1 二次风挡板	A侧炉烟氧量
层2 二次风挡板	A侧空预器出口处烟温
层4 二次风挡板	B侧二次风箱与炉膛差压
层6 二次风挡板	B侧二次风箱入口压力
层8 二次风挡板	B侧炉膛出口温度
层10 二次风挡板	B侧炉烟氧量
层12 二次风挡板	B侧空预器出口处烟温

### 3 模型构建过程

图2为模型构建流程。首先对原始数据集进行

初步数据处理;然后利用基于随机森林模型的RFE方法提取出重要特征作为模型的输入特征;接着随机将90%的数据作为训练集,10%的数据作为测试集,建模比较XGBoost、随机森林等6种机器学习模型预测效果,并进行5次-10折交叉检验以保证模型输出结果的稳健性;最后运用均方误差(MSE)等多种模型评价指标展现各模型预测效果。

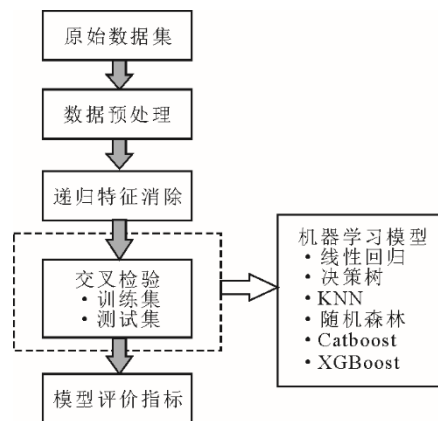


图2 建模流程  
Fig.2 Modeling process

#### 3.1 数据预处理

由于飞灰含碳量数据采样时间间隔较大,为了能更科学、准确地将采样间隔为8 h的飞灰含碳量数据与采样间隔为1 min的输入特征相对应,运用负荷率对飞灰含碳量进行调整。

通常当锅炉处于低负荷和高负荷时,产生的飞灰含碳量较高;当锅炉处于中高负荷水平运转时,产生的飞灰含碳量较少。根据经验,假设锅炉负荷率50%为低负荷,该状态下每分钟飞灰含碳量为每8 h平均飞灰含碳量的1~1.3倍;负荷率75%为中高负荷,该状态下每分钟飞灰含碳量为每8 h平均飞灰含碳量的0.75~1倍;负荷率100%为高负荷,该状态下每分钟飞灰含碳量为每8 h平均飞灰含碳量的0.9~1.1倍。利用低中高负荷下假设的飞灰含碳量与8 h平均飞灰含碳的关系,拟合锅炉负荷率和实时飞灰含碳量间的关系,如图3所示。

在低、中、高负荷中,将实时飞灰含碳量与8 h平均飞灰含碳量的比值范围平均划分,共取1 000个测试样本放入随机森林模型中进行检验,发现当低、中、高负荷下的调整倍数分别为1.03、0.80和1.00时,随机森林模型取得了最好的预测效果,最终得到拟合公式为:

$$Y = (2.78L^2 - 5.22L + 3.44)Y \quad (1)$$

式中： $Y$  为由拟合公式计算出的飞灰含碳量，%，频率为 1 min/次； $L$  为负荷率，%，取机组负荷 330 MW； $Y$  为实测飞灰含碳量，%，取 8 h 平均值。

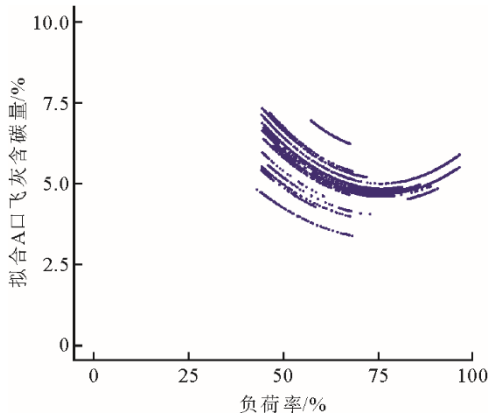


图 3 负荷率与拟合飞灰含碳量间的关系

Fig.3 The relationship between the load rate and the fitted carbon content of fly ash

图 4 为实测飞灰含碳量与拟合公式计算出的飞灰含碳量之间的关系。由图 4 可知，拟合值与真实值差距在 1.5% 以内。

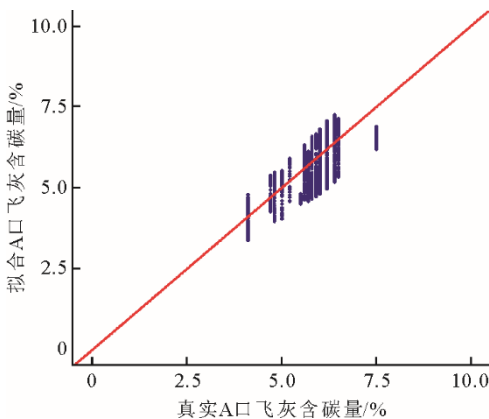


图 4 真实飞灰含碳量和拟合飞灰含碳量的关系

Fig.4 The relationship between real carbon content of fly ash and the fitted one

机组负荷和主蒸汽流量间的关系如图 5 所示。由图 5 可知，锅炉的主蒸汽流量和机组负荷相关系数高达 0.98。为此，将数据按主蒸汽流量分为低、中、高负荷 3 组（图 6），其中低负荷组主蒸汽流量在 600 t/h 以下，中负荷组主蒸汽流量 600~800 t/h，高负荷组主蒸汽流量在 800 t/h 以上。后续将在这 3 种工况条件下分别预测锅炉飞灰含碳量。

### 3.2 特征选择

通过相关性分析（图 7）发现主蒸汽流量、机组负荷和总风量间存在较强相关性，两两相关系数在 0.98 以上。由于负荷率已被用于分配飞灰含碳

量，因此这里先删除机组负荷特征避免人为影响。

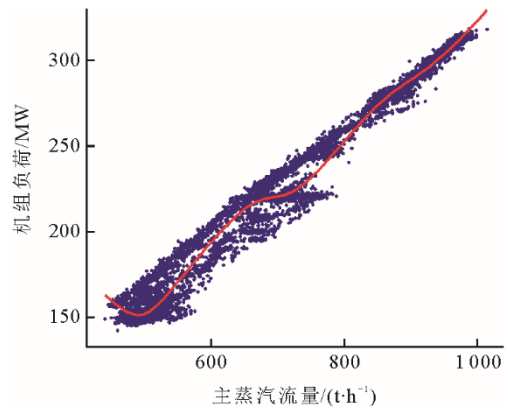


图 5 机组负荷和主蒸汽流量间的关系

Fig.5 The relationship between the load and main steam flow

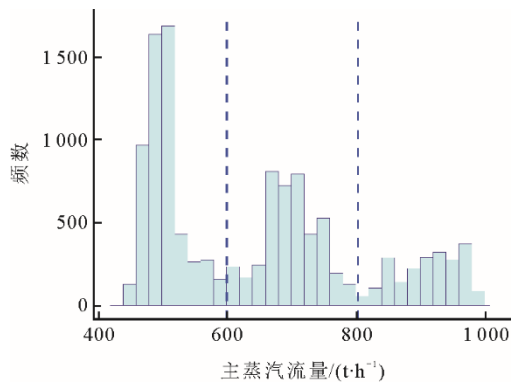


图 6 主蒸汽流量频数分布

Fig.6 Frequency distribution of main steam flow

本文采用了线性回归、装袋树和随机森林 3 种模型进行递归特征消除，3 种模型在不同特征个数下的预测结果指标对比如图 8 所示。由图 8 可知，采用随机森林的拟合效果最好。在随机森林下，选择 RMSE 和  $R^2$  综合最优的 9 个特征作为模型最终的输入特征，这 9 个特征分别是层 3 给粉机转速之和、总风量、A 侧二次风箱与炉膛差压、层 2 二次风挡板、层 4 二次风挡板、A 侧炉烟氧量、一次风管与炉膛差压、燃烧器摆角指令信号 A、A 侧空预器出口处烟温。

### 3.3 模型评价指标和参数的选择

在评价指标的选取上，本文不仅采用了机器学习常用评价指标均方误差 (MSE)，同时采用了预测值与真实值偏差小于 0.2 和 0.4，相对偏差小于 5% 和 10% 2 类评价指标，具体见表 3。其中 MSE 越小，代表模型预测效果越好，与真实值的偏差越小，模型预测效果越好。

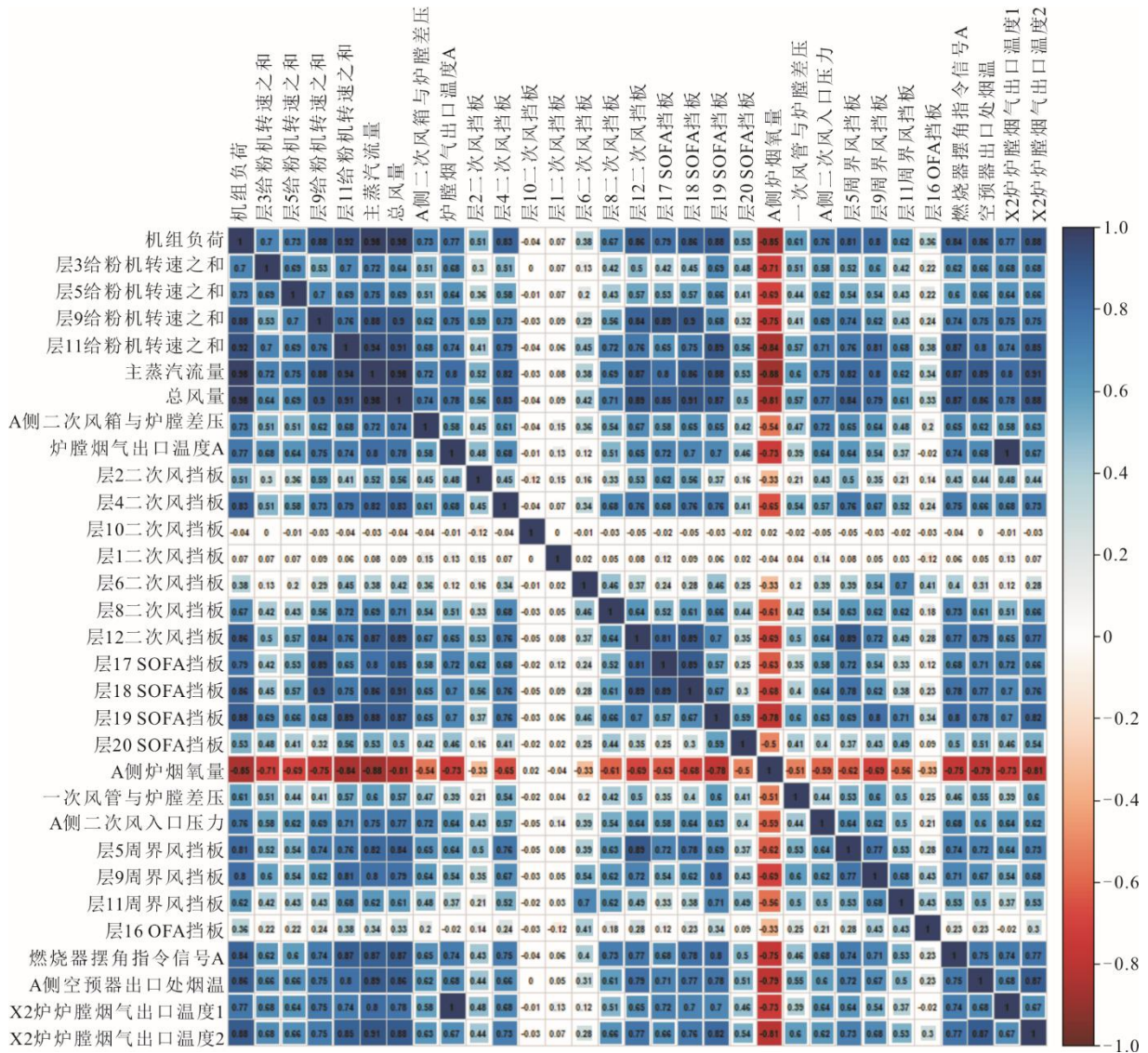


图7 变量的相关系数  
Fig.7 Correlation coefficients of the variables

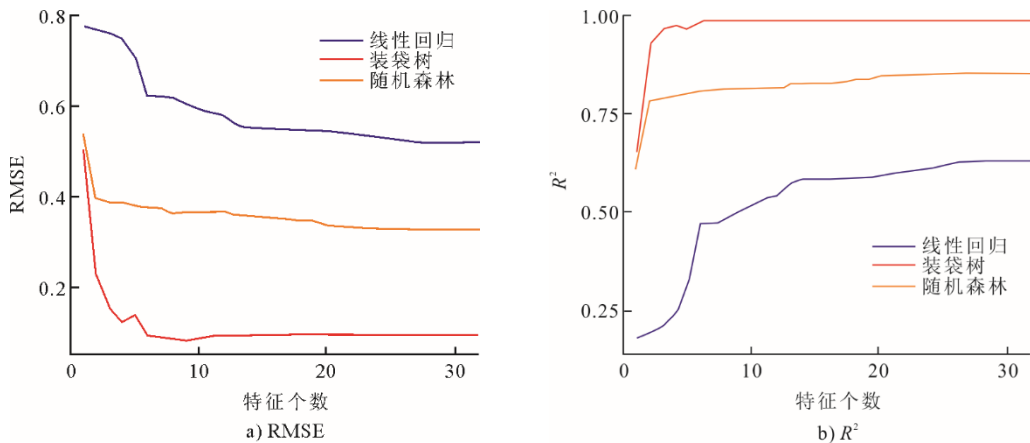


图8 递归特征消除下3种模型在不同特征个数下的预测结果指标对比

Fig.8 Comparison of prediction results of three models with different number of features based on recursive feature elimination

表 3 模型评价指标  
Tab.3 Model evaluation indicators

评价指标	公式	说明
MSE	$\delta_{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$	真实值与预测值之差的平均值
偏离 0.2	$P_{er0.2} = \frac{1}{n} \sum_{i=1}^n ( Y_i - \hat{Y}_i  < 0.2)$	预测值偏离真实值±0.2
偏离 0.4	$P_{er0.4} = \frac{1}{n} \sum_{i=1}^n ( Y_i - \hat{Y}_i  < 0.4)$	预测值偏离真实值±0.4
偏离 5%	$P_{er5\%} = \frac{1}{n} \sum_{i=1}^n \left( \left  \frac{Y_i - \hat{Y}_i}{Y_i} \right  < 5\% \right)$	预测值偏离真实值±5%
偏离 10%	$P_{er10\%} = \frac{1}{n} \sum_{i=1}^n \left( \left  \frac{Y_i - \hat{Y}_i}{Y_i} \right  < 10\% \right)$	预测值偏离真实值±10%

在模型参数的选择上，对 KNN 模型，通过调整 KNN 临近点个数以提高模型的准确度，具体如图 9 所示。由图 9 可知，选择临近点个数为 3 时 KNN 模型预测效果较好。对于随机森林模型，选择 70 棵树时模型预测效果较好（图 10）。

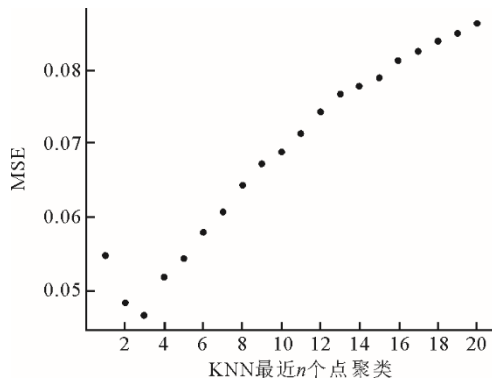


图 9 KNN 模型聚类点  
Fig.9 KNN model clustering points

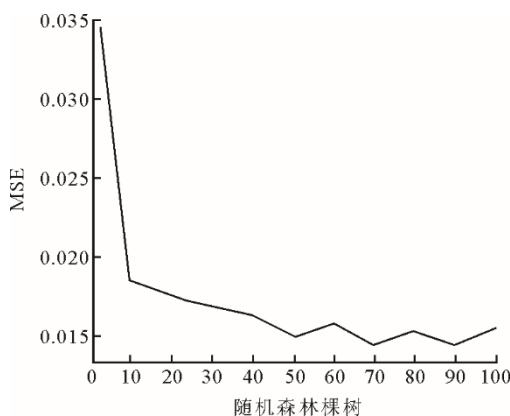


图 10 随机森林模型棵树  
Fig.10 The number of trees of random forest model

对多个超参数，采用网格搜索方法进行确定。决策树模型有 minsplit、cp、maxdepth 3 个重要的超参数，minsplit 参数存放了在被迫创建一个叶节点之前，算法尝试一次分裂所需数据点的最小数量；cp 为复杂度参数，默认值为 0.01；maxdepth 参数是叶节点和根节点之间的允许的节点数量最大值。根据经验，本文将 minsplit 参数调节范围设定在 5~30，间隔为 5，cp 参数范围设置在 0~0.5，间隔为 0.05，maxdepth 参数范围设置在 1~20，间隔为 1，测试了在这 3 个参数的组合情况下模型拟合效果，最终得到在 minsplit=10、cp=0、maxdepth=17 时模型拟合效果最好，MSE 为 0.048。

XGBoost 的超参数主要有 eta 和 max\_depth。eta 参数是学习率，默认值为 0.3，较小的学习率可以让后面的学习更加仔细；max\_depth 参数是每颗树的最大深度，默认值为 6。本文将最大迭代次数设定为 200，eta 的调参范围设置为 0~1，间隔为 0.05，max\_depth 的调参范围设置在 1~20，间隔为 1。根据测试情况，XGBoost 在 eta=0.15、max\_depth=8 时数据预测效果最好，MSE 为 0.017。

### 4 结果与分析

分别使用简单线性回归模型、决策树模型、KNN 聚类模型、随机森林模型、Catboost 模型和 XGBoost 模型在测试集上对 A 侧飞灰含碳量进行预测。结果发现，XGBoost 模型预测效果最好（图 11），随机森林模型次之（图 12），但预测精确度与 XGBoost 相差较小。决策树模型和 KNN 模型预测效果整体不如 XGBoost 模型和随机森林模型，而简单线性模型和 Catboost 模型预测结果较真实值偏离较大。

表 4—表 6 为多模型评价指标情况。由表 4—表 6 可见，XGBoost 模型、随机森林模型、决策树模型、KNN 模型的预测值在偏离真实值 5% 范围内比例超过 98%，在偏离真实值 10% 范围内比例超过 99%，预测准确性较高。其中 XGBoost 模型预测效果最好，综合预测精度最高，且训练集和测试集的误差值差距较小。B 侧飞灰含碳量和 AB 侧平均飞灰含碳量的预测结果与 A 侧相似，均为 XGBoost 和随机森林模型预测表现最好。

6 种模型下的残差分布如图 13 所示。由图 13 可知：决策树模型、KNN 模型、随机森林模型和 XGBoost 模型的残差分布较线性回归模型和 Catboost 模型更为集中；XGBoost 模型的残差分布集中程度最高；线性回归模型和 Catboost 模型残差

分散程度高,在实际运用时可能产生较大的预测偏差;决策树模型、KNN 模型、随机森林模型和

XGBoost 模型残差集中于 0 附近,且呈左右对称分布,故更适合预测飞灰含碳量。

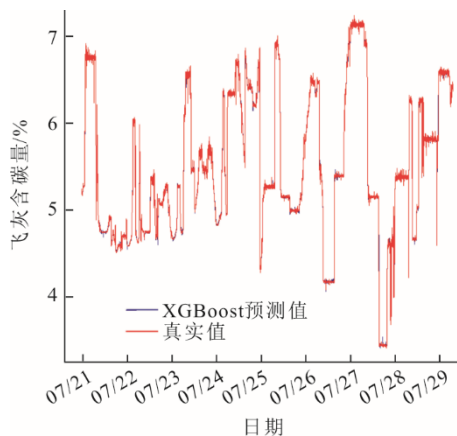


图 11 XGBoost 模型预测结果  
Fig.11 Prediction result of XGBoost model

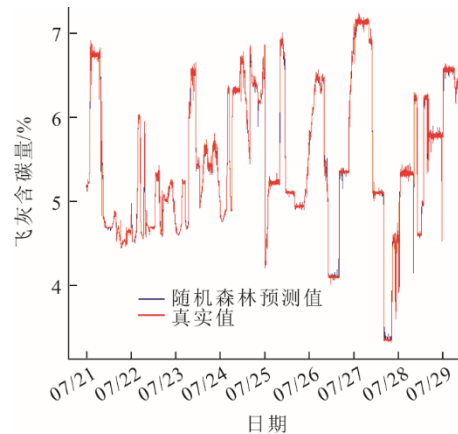


图 12 随机森林模型预测结果  
Fig.12 Prediction result of random forest model

表 4 6 种机器学习模型 A 侧飞灰含碳量预测结果  
Tab.4 Prediction results of the six machine learning models on the carbon content of fly ash of port A

学习模型	测试集/训练集	MSE	偏离真实值 5%/%	偏离真实值 10%/%	偏离真实值 0.2%/	偏离真实值 0.4%/
线性回归	训练集	0.336	40.53	66.64	30.78	55.72
	测试集	0.336	40.53	66.61	30.73	55.69
决策树	训练集	0.003	99.46	99.82	99.12	99.69
	测试集	0.010	98.41	99.37	97.56	99.01
KNN	训练集	0.002	99.73	99.90	99.55	99.80
	测试集	0.009	98.84	99.47	98.21	99.17
随机森林	训练集	0.003	99.42	99.83	99.08	99.37
	测试集	0.006	98.76	99.64	98.07	99.37
Catboost	训练集	0.144	66.52	87.54	53.86	79.90
	测试集	0.299	57.75	78.64	45.05	70.65
XGBoost	训练集	0.002	99.71	99.89	99.52	99.80
	测试集	0.006	98.97	99.69	98.29	99.49

表 5 6 种机器学习模型 B 侧飞灰含碳量预测结果  
Tab.5 Prediction result of the six machine learning models on the carbon content of fly ash of port B

学习模型	测试集/训练集	MSE	偏离真实值 5%/%	偏离真实值 10%/%	偏离真实值 0.2%/	偏离真实值 0.4%/
线性回归	训练集	2.742	2.05	5.38	2.16	5.35
	测试集	2.745	2.05	5.38	2.16	5.36
决策树	训练集	0.026	98.29	98.82	98.26	98.81
	测试集	0.069	97.02	97.82	96.85	97.85
KNN	训练集	0.016	99.12	99.39	99.12	99.42
	测试集	0.049	98.50	98.85	98.35	98.88
随机森林	训练集	0.024	97.00	98.45	96.97	98.54
	测试集	0.040	94.73	97.34	94.61	97.42
Catboost	训练集	0.561	27.80	51.27	27.47	51.61
	测试集	0.955	25.90	47.45	26.05	47.92
XGBoost	训练集	0.016	99.12	99.39	99.12	99.42
	测试集	0.041	97.28	98.38	97.12	98.31

表 6 6 种机器学习模型 AB 侧平均飞灰含碳量预测结果  
 Tab.6 Prediction result of the six machine learning models on the carbon content of fly ash of average A & B

学习模型	测试集/训练集	MSE	偏离真实值 5%/%	偏离真实值 10%/%	偏离真实值 0.2%/	偏离真实值 0.4%/
线性回归	训练集	3.302	14.31	34.63	5.18	10.51
	测试集	3.308	14.29	34.58	5.19	10.51
决策树	训练集	0.034	98.92	99.45	98.44	98.81
	测试集	0.088	98.16	98.77	97.53	98.02
KNN	训练集	0.024	99.40	99.66	99.20	99.34
	测试集	0.076	98.74	99.04	98.50	98.67
随机森林	训练集	0.035	98.61	99.44	96.41	98.14
	测试集	0.059	97.63	99.09	94.07	96.82
Catboost	训练集	0.561	27.80	51.27	27.47	51.61
	测试集	0.955	25.90	47.45	26.05	47.92
XGBoost	训练集	0.024	99.39	99.65	99.21	99.34
	测试集	0.052	98.43	99.23	96.95	98.08

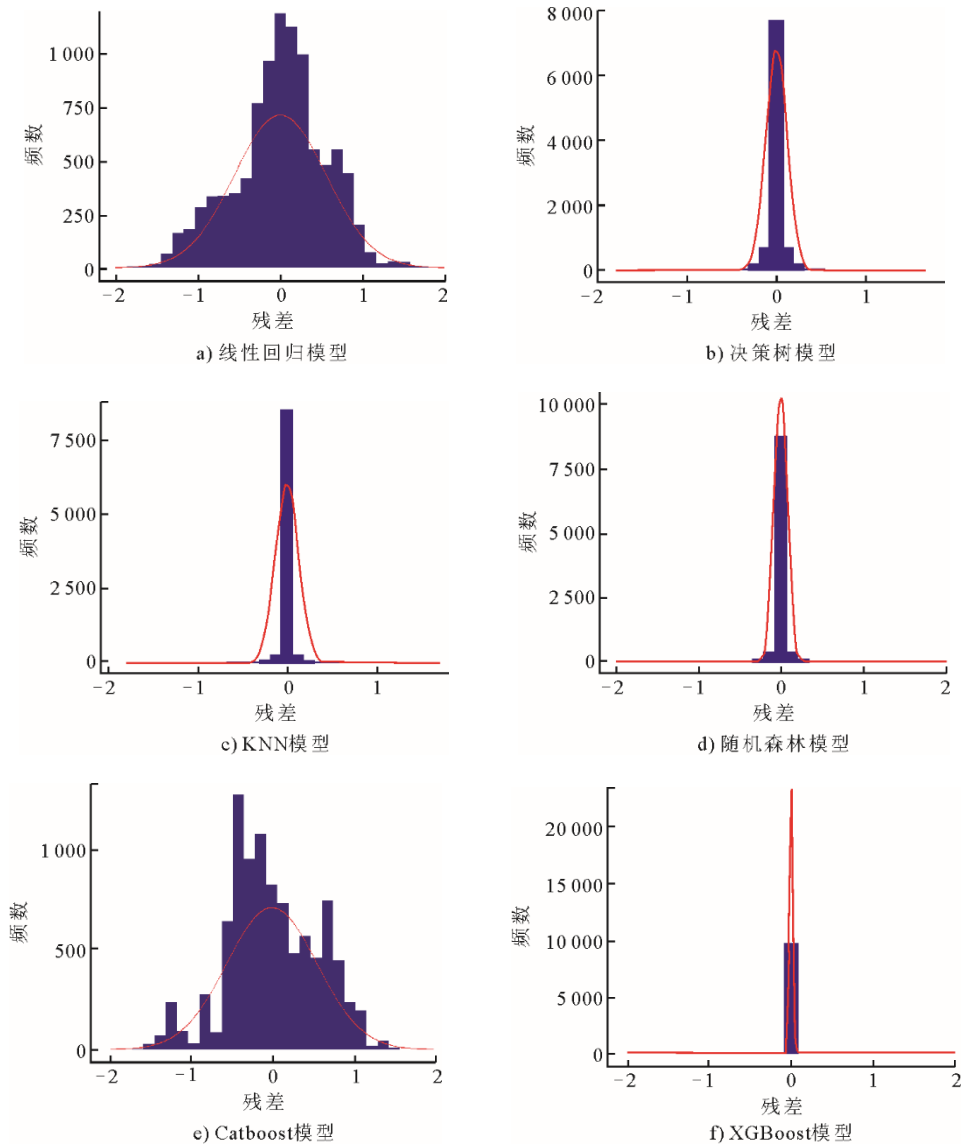


图 13 6 种机器学习模型残差分布  
 Fig.13 Residual distribution plot of the six machine learning models

表7为各模型训练和预测时长数据。其中,训练时长为各模型经过训练及5次-10折验证后输出稳健预测结果的时间,预测时长为各模型预测测试集结果所需要的时间。由表7可见:线性回归模型、决策树模型和KNN模型的训练时长明显短于随机森林模型、Catboost模型和XGBoost模型;在预测时长上,KNN模型预测时间最长,其次是随机森林模型,线性回归模型最短;决策树、KNN、随机森林和XGBoost 4个预测效果较好的模型中XGBoost模型的预测速度最快。

表7 6种机器学习模型训练与预测时长 单位:s  
Tab.7 Training and prediction time of the six machine learning models

学习模型	线性回归	决策树	KNN	随机森林	Catboost	XGBoost
训练时长	0.446	20.300	27.611	386.682	283.614	111.216
预测时长	0.000 6	0.006 4	0.102 1	0.019 9	0.002 6	0.003 4

利用决策树、随机森林模型进行了特征重要性排序(图14、图15)。由图14和图15可知,总风量、燃烧器摆角指令信号A、A侧空预器出口处烟温是较为重要的特征,在决策树和随机森林的特征重要性排序中都位居前列。层3给粉机转速之和和一次风管与炉膛差压的重要性排序相对靠后,说明这2个特征对飞灰含碳量的影响较小。

以随机森林模型为例进行分工况实验,预测结果见表8。由表8可以看出:随机森林模型在低、中、高负荷水平下均有很好的预测效果,3种工况

表8 不同负荷下随机森林模型预测结果  
Tab.8 Random forest prediction results based on different loads

工况	测试集/训练集	MSE	偏离真实值 5%/%	偏离真实值 10%/%	偏离真实值 0.2%/	偏离真实值 0.4%/
低负荷	训练集	0.003	99.45	99.95	98.88	99.78
	测试集	0.006	98.61	99.66	97.42	99.25
中负荷	训练集	0.005	99.03	99.63	98.86	99.47
	测试集	0.009	98.08	99.25	97.66	99.09
高负荷	训练集	0.000 6	99.93	100.00	99.84	100.00
	测试集	0.001 4	99.80	99.95	99.64	99.95

## 5 结 论

本文以四角切圆燃烧煤粉锅炉的运行数据为基础,利用锅炉负荷率将飞灰含碳量的数据频率调整为每分钟一次,与输入特征采样间隔相一致,同时利用基于随机森林的递归特征消除方法提取出主要特征,建立了线性回归、决策树、KNN、随机

下模型预测飞灰含碳量偏离真实值 $\pm 5\%$ 的准确度均高于98%;其中,高负荷水平下模型表现最佳,训练集和测试集下的MSE分别为0.000 6和0.001 4,显著低于低负荷和中负荷,高负荷下其余4个评价指标也均高于99%,预测十分准确。

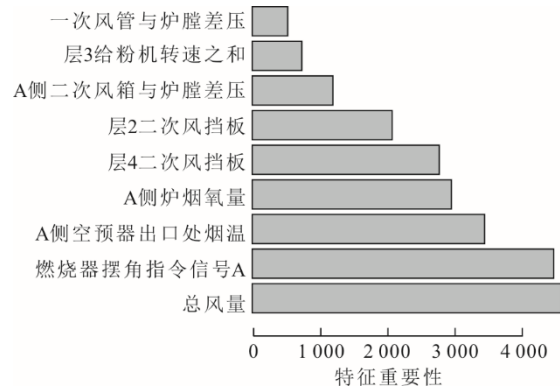


图14 决策树模型下的特征重要性排序  
Fig.14 Feature importance ranking in decision tree

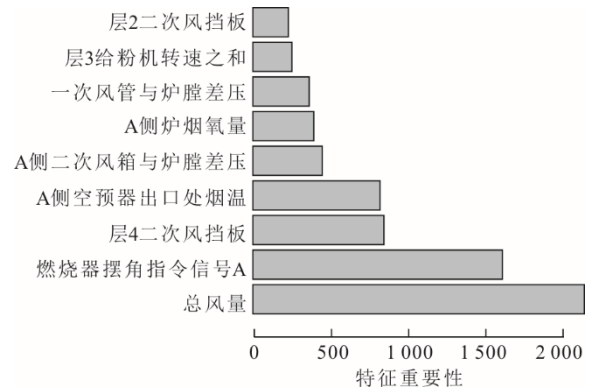


图15 随机森林模型下的特征重要性排序  
Fig.15 Feature importance ranking in random forest

森林、Catboost和XGBoost模型来预测锅炉飞灰含碳量以及及时了解锅炉飞灰含碳量的变化趋势。研究结果表明如下。

1) 基于随机森林的递归特征消除方法可将初始的30个特征提取出最重要的9个特征,利用这9个特征就可反映出飞灰含碳量的大部分信息,在提高模型计算速度的同时也提高了模型的准确度。

2) 决策树模型在训练集上的均方误差为 0.003, 测试集上均方误差为 0.010; KNN 模型在训练集上的均方误差为 0.002, 测试集上均方误差为 0.009; 随机森林模型在训练集上的均方误差为 0.003, 测试集上均方误差为 0.006; XGBoost 模型在训练集上的均方误差为 0.002, 测试集上均方误差为 0.006。这 4 种模型均明显优于线性回归 (测试集均方误差 0.336) 和 Catboost (测试集均方误差 0.299) 2 种模型, 其中 XGBoost 综合预测精度最高。

3) 总风量、燃烧器摆角指令信号 A、A 侧空预器出口处烟温是对飞灰含碳量影响较大的特征, 层 3 给粉机转速之和和一次风管与炉膛差压对飞灰含碳量的影响相对较小。

4) 由于 XGBoost 模型综合预测效果最好, 预测时间较短, 计算速度较快, 本文推荐使用这 6 个模型中的 XGBoost 模型对飞灰含碳量进行预测。

#### [参考文献]

[1] 王春林, 周昊, 周樟华, 等. 基于支持向量机的大型电厂锅炉飞灰含碳量建模[J]. 中国电机工程学报, 2005, 25(20): 72-76.  
WANG Chunlin, ZHOU Hao, ZHOU Zhanghua, et al. Support vector machine modeling on the unburned carbon in fly ash[J]. Proceedings of the CSEE, 2005, 25(20): 72-76.

[2] 陈敏生, 刘定平. 基于核主元分析和支持向量机的电站锅炉飞灰含碳量软测量建模[J]. 华北电力大学学报, 2006(1): 72-75.  
CHEN Minsheng, LIU Dingping. Soft-sensing modeling of the unburned carbon in fly ash based on KPCA-SVM for power station boilers[J]. Journal of North China Electric Power University, 2006, 33(1): 72-75.

[3] 周国雄, 李琳, 沈学杰. 基于 SVM 和灰色预测的飞灰含碳量集成预测[J]. 系统仿真学报, 2013, 25(4): 727-731.  
ZHOU Guoxiong, LI Lin, SHEN Xuejie. Integrated prediction for carbon content in fly ash based on online support vector machine and grey prediction[J]. Journal of System Simulation, 2013, 25(4): 727-731.

[4] 卞和营, 方彦军. 基于支持向量回归的飞灰含碳量软测量[J]. 热力发电, 2014, 43(10):46-50.  
BIAN Heyin, FANG Yanjun. SVR-based study on soft-sensing measurement of carbon content in fly ash[J]. Thermal Power Generation, 2014, 43(10): 46-50.

[5] 周昊, 朱洪波, 曾庭华, 等. 基于人工神经网络的大型电厂锅炉飞灰含碳量建模[J]. 中国电机工程学报, 2002, 22(6): 96-100.  
ZHOU Hao, ZHU Hongbo, ZENG Tinghua, et al. Artificial neural network modelling on the unburned carbon in fly ash from utility boilers[J]. Proceedings of the CSEE, 2002, 22(6): 96-100.

[6] 方湘涛, 叶念渝. 基于 BP 神经网络的电厂锅炉飞灰含碳量预测[J]. 华中科技大学学报(自然科学版), 2003, 31(12): 75-77.  
FANG Xiangtao, YE Nianyu. A system for forecasting the unburned carbon of the fly ash from utility boilers based on BP artificial neural networks[J]. Journal of Huazhong University of Science and Technology(Natural Science

Edition), 2003, 31(12): 75-77.

[7] 赵新木, 王承亮, 吕俊复, 等. 基于 BP 神经网络的煤粉锅炉飞灰含碳量研究[J]. 热能动力工程, 2005, 20(2): 158-162.  
ZHAO Xinmu, WANG Chengliang, LYU Junfu, et al. The investigation of carbon content in fly ash for a BP neural network-based pulverized coal-fired boiler[J]. Journal of Engineering for Thermal Energy and Power, 2005, 20(2): 158-162.

[8] 温文杰, 马晓茜, 刘翱. 锅炉混煤掺烧的飞灰含碳量预测与运行优化[J]. 热力发电, 2010, 39(9): 30-35.  
WEN Wenjie, MA Xiaoqian, LIU Ao. Prediction of unburned carbon content in fly ash and operation optimization for mixedly burning blended coal in boilers[J]. Thermal Power Generation, 2010, 39(9): 30-35.

[9] 叶兆青, 鄢晓忠, 卢坤杰, 等. 基于 BP 神经网络的超临界对冲火焰锅炉飞灰含碳量预测分析[J]. 锅炉制造, 2014(5): 1-10.  
YE Zhaoqing, YAN Xiaozhong, LU Kunjie, et al. A predictive analysis of carbon content in fly ash based on BP neural network[J]. Boiler Manufacturing, 2014(5): 1-10.

[10] 冯旭刚, 钱家俊, 章家岩. 基于遗传神经网络敏感度分析的飞灰含碳量测量方法[J]. 电子测量与仪器学报, 2016, 30(7): 1083-1089.  
FENG Xugang, QIAN Jiajun, ZHANG Jiayan. Prediction method of unburned carbon content in fly ash based on genetic neural network with sensitivity analysis[J]. Journal of Electronic Measurement and Instrumentation, 2016, 30(7): 1083-1089.

[11] 乔源, 王建峰, 杨永存, 等. 基于神经网络的飞灰含碳量软测量模型及实现[J]. 电力科学与工程, 2019, 35(11): 55-61.  
QIAO Yuan, WANG Jianfeng, YANG Yongcun, et al. Soft measurement model and implementation of fly ash carbon content based on neural network[J]. Electric Power Science and Engineering, 2019, 35(11): 55-61.

[12] 王芳, 马素霞, 王河. 基于随机森林变量选择的飞灰含碳量预测模型[J]. 热力发电, 2018, 47(11): 89-95.  
WANG Fang, MA Suxia, WANG He. Prediction model of carbon content in fly ash using random forest variable selection method[J]. Thermal Power Generation, 2018, 47(11): 89-95.

[13] 朱璉琦, 牛晓凡, 肖显斌. 基于改良 BP 神经网络的生物质锅炉飞灰含碳量预测模型研究[J]. 可再生能源, 2020, 38(2): 150-157.  
ZHU Jinqi, NIU Xiaofan, XIAO Xianbin. Prediction models of the carbon content of fly ash in a biomass boiler based on improved BP neural networks[J]. Renewable Energy Resources, 2020, 38(2): 150-157.

[14] 张先炼, 肖禹, 潘波, 等. 基于分层多项式模型的飞灰含碳量预测[J]. 能源与节能, 2021(12): 166-169.  
ZHANG Xianlian, XIAO Yu, PAN Bo, et al. Prediction of carbon content in fly ash based on hierarchical and polynomial model[J]. Energy and Energy Conservation, 2021(12): 166-169.

[15] 吴辰文, 梁靖涵, 王伟, 等. 基于递归特征消除方法的随机森林算法[J]. 统计与决策, 2017(21): 60-63.  
WU Chenwen, LIANG Jinghan, WANG Wei, et al. Random forest algorithm based on recursive feature elimination[J]. Statistics & Decision, 2017(21): 60-63.

(责任编辑 邓玲惠)